A Neural Map View of Planetary Spectral Images for Precision Data Mining

and Rapid Resource Identification

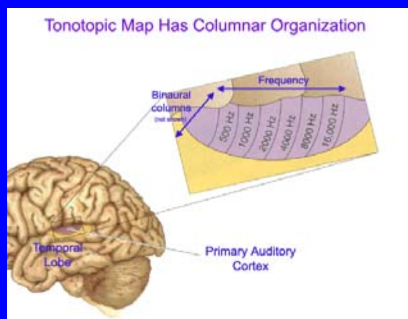
Erzsébet Merényi, PI, Rice University, Houston, TX
William H. Farrand, Co-I, Space Science Institute, Boulder, CO
Robert H. Brown, Co-I, University of Arizona, Tucson, AZ
Thomas Villmann, Collaborator, University of Leipzig, Germany
Colin Fyfe, Collaborator, University of Paisley, Scotland

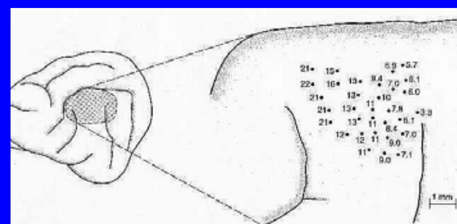
Additional Contributors:
Abha Jain, Kadim Tasdemir, Mike Mendnhall, Graduate Students, Rice University
Philip Tracadas, Research Programmer, Rice University
Allen Geer, Undergraduate Student, Rice University

Support by AISRP NASA OSSA / Science Directorate
Project web page: www.ece.rice.edu/~erzsebet/HYPEREYE.html

Neural Maps in the Brain

Example: In the auditory cortex *tonotopic maps* are formed where the 2-D spatial order of cell responses corresponds to the acoustic frequency of tones perceived. This spatial organization according to similarities facilitates precise and fast recognition and retreival of patterns. This is the type of learning that interests us.





A close-up of the cat's auditory receiving area (shaded), showing the tonotopic map at the cortical surface. Each large dot represents a single neuron. The number beside each dot is the characteristic frequency of that neuron in 1000 Hz. *(From Abeles and Goldstein, 1970, and Gidick, Gescheider and Frisina, 1989.)*

Topology preserving mapping (learning) of sensory stimuli on the cortex (2-D surface).

Left figure:

A cartoon of the spatial ordering of sensory (auditory) stimuli, according to their similarities, on the cortex.

Right figure:

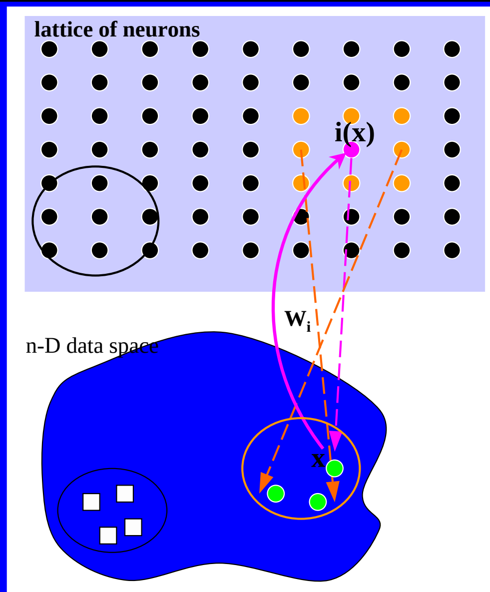
The cat's brain, indicating the location of the auditory receiving area (shaded), and a close-up of the auditory receiving area, showing the tonotopic map at the cortical surface. Each large dot represents a single neuron, and the number beside each dot represents the characteristic frequency of that neuron in thousands of Hz. (From Abeles and Goldstein, 1970, and Gidick, Gescheider and Frisina, 1989.)

Note the ordering of frequencies in both directions, i.e., finer details within each column of major frequency ranges.

This type of topology preserving mapping (learning) of sensory stimuli on a 2-D (low-D) surface is of our interest. Other biological analogs include the retinotopic maps in the visual cortex, or stimuli percieved through touching organized in the somatosensory cortex.

Self-Organizing Neural Maps

(unsupervised learning machine)



Formation of basic (Kohonen) SOM:

$x = (x1, x2, \ldots, xn) \in M \subseteq R^n$ input pattern

$w_j = (w_{j1}, w_{j2}, \ldots, w_{jn})$ j=1, … , N synaptic weight vector (pointer)

3-step learning, through many iterations:

1. Competition

Select a pattern x randomly. Winning neuron $i(x) = \arg\min_j \|x - w_j\|$, j=1, … , N

2. Cooperation: Winner activates neighbor neurons according to a neighborhood function $h_{j,i(x)}(t)$.

3. Synaptic adaptation

$w_j(t+1) = w_j(t) + a(t)\, h_{j,i(x)}(t)\,(x - w_j(t))$

During learning, areas of neurons form to collectively represent groups of similar patterns ⇒ *clustering*

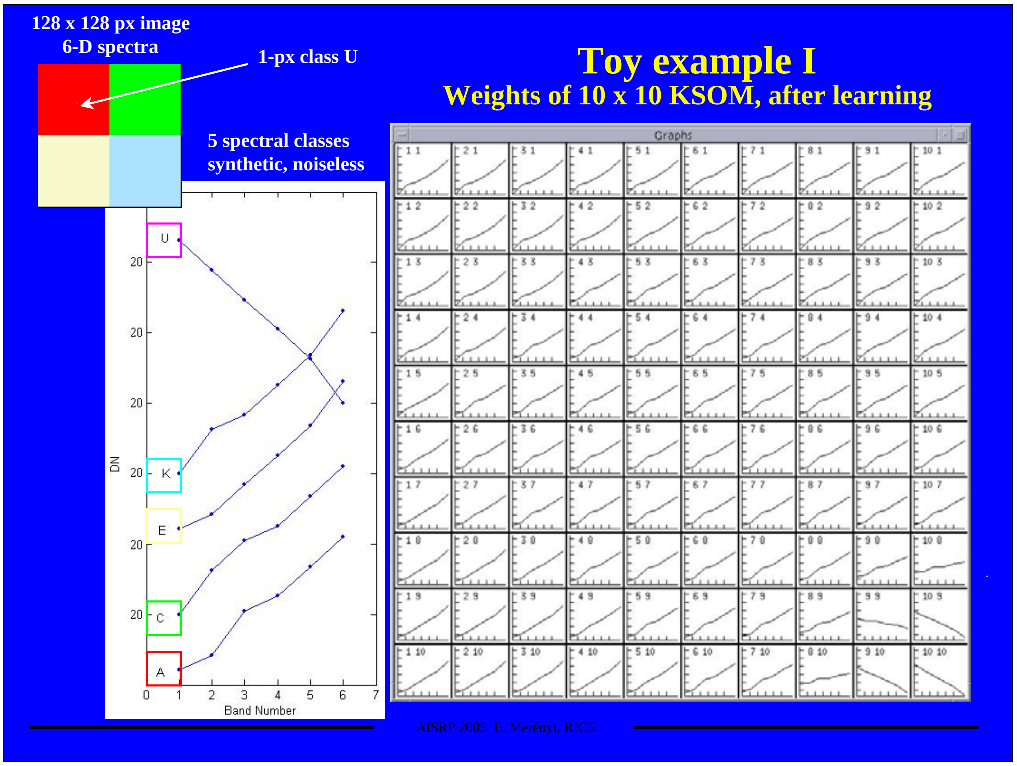
- neighborhood preserving, adaptive vector quantizer (optimal placement of prototypes)
- ***nonlinear*** mapping of the n-D input space to a low-D lattice
- measure of dissimilarities is expressed by the difference of the weights of neighbor neurons

The artificial machine learning model that intends to mimic the cortical ordering of stimuli is the Self-Organizing Map.

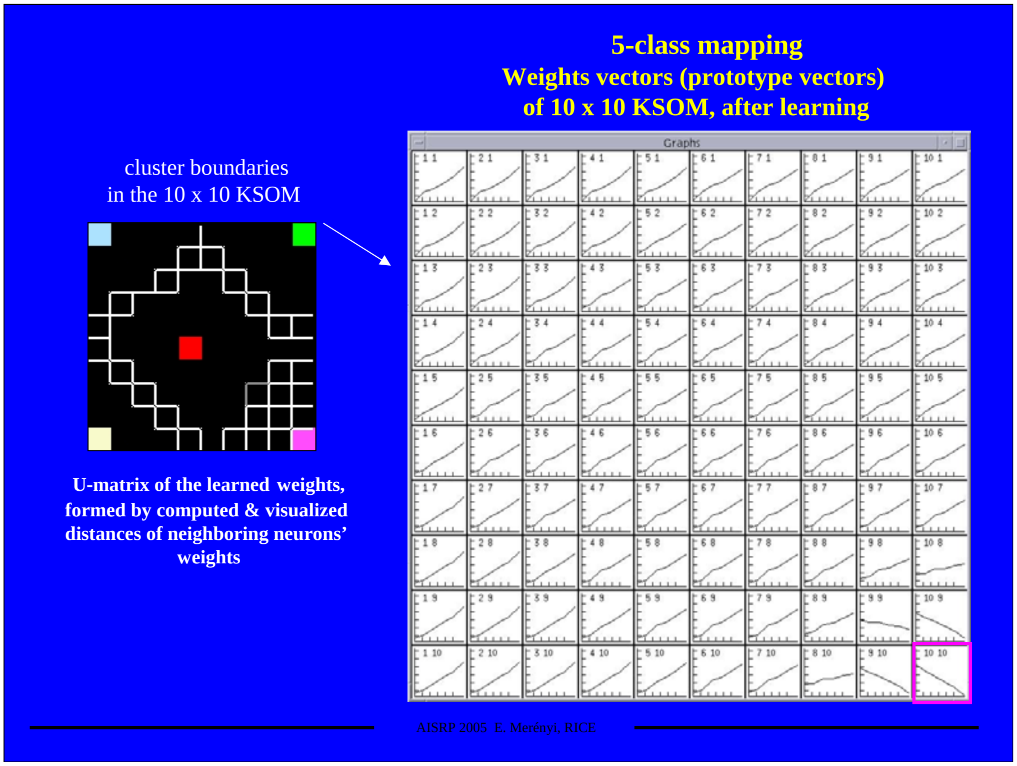
This machine learns the structure of the input data by cycling through the above steps 1 – 3, many times.

Important is the neighborhood preservation (topologically faithful mapping) … Without neighborhood concept in the SOM lattice, this would be ordinary vector quantization.

Main reference: Kohonen Self-Organizing Maps, (Springer-Verlag, 1997, 1999).



An illustration of learning the data structure by an SOM, through a small spectral image where each pixel is a 6-D spectrum. There are four large classes and one tiny - 1-pixel – class in this image. The mean of the class spectra are shown in the left plot. On the right, the learned SOM weight vectors are shown (as spectra) at their corresponding locations in the 10 x 10 SOM grid. It is clear that the weight in the lower right corner became the prototype of the tiny class U, the weights in the middle represent class A, those in the upper left, lower left corner, and the upper right corners represent K, E, and C, respectively. Note the relative size of the SOM areas (the number of neurons) allocated for each class. The large classes get much larger areal representation than the tiny class.



By computing and visualizing the distances between weights of neighbor neurons cluster boundaries are outlined as shown on the left. Colors within the SOM areas separated by the white fences here show the class types known to us. The cluster boundaries represent what the SOM learned. Clearly, the agreement is perfect. Highlighting in the input data space (the image) all data points that were mapped to neurons within the clusters detected by the SOM produces perfect coverage of the respective known class areas.

Note that the 1-pixel class is only represented by 1 weight vector (the other two that look similar on this scale are dissimilar enough to be "fenced off" on the left image).

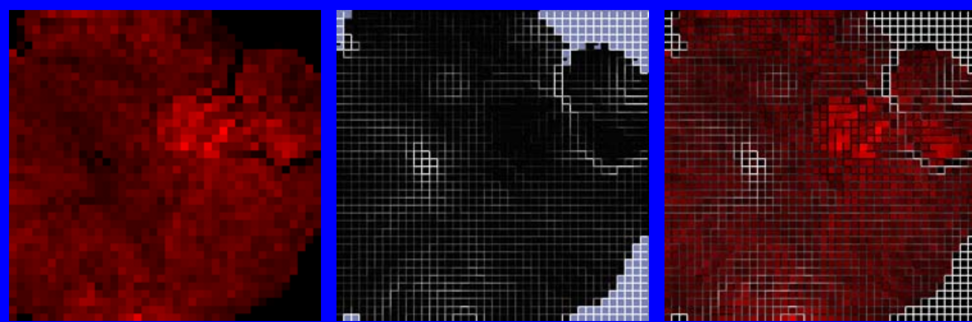
Note also that this image is noiseless, syntethic, all spectra are exactly the same within each class, so at the end all spectra from a class map onto the same one neuron. However, the learned weights occupy the entire area for a class within the cluster boundaries.

Possible representations of SOM knowledge for cluster detection

(“remap” HYPEREYE module)

Input data: AVIRIS image of Lunar Crater Volcanic Field, 420 x 614 pixels x 194 bands

Density Map | Fence Map (U-matrix) | Combined Density & Fence Map



For real, high-dimensional data, the SOM shows fuzzy, complicated structure. Density and fences together represent the full knowledge of the SOM.

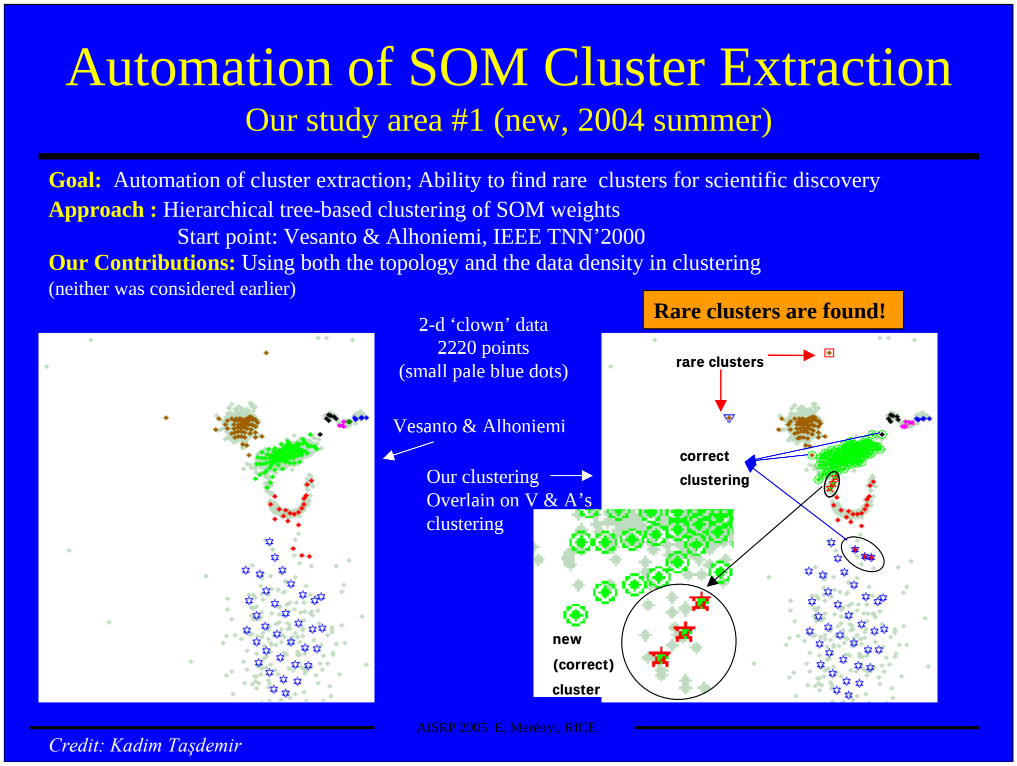
Automation of cluster extraction from SOMs of this complexity is an unsolved challenge.

Real, noisy, high-dimensional data produce fuzzy, complicated SOM knowledge structure. Capturing the clusters correctly and to a desired granularity is a great challenge. Automation of cluster extraction from a learned SOM is currently an unsolved problem for this level of complexity.

We are working on various issues related to the interpretation of maps like these. Ideally, the combined density and fence map provides the full knowledge of the SOM. However, many small details need to be addressed. For example, by overlaying the two maps the fences around the brightest group became almost invisible due to contrast and color relations that trick the human eye.

We have built in various on-the-fly capabilities to change such visualization parameters instantaneously during a “remap” session. This is important also because optimal display for the human eye may vary from one part of the SOM to another.

Insight gained by using these semi-manual tools helps get closer to full automation of the interpretation of the SOM.



Goal: Automation of cluster extraction with desired granularity; Ability to find rare clusters for scientific discovery

Approach : Hierarchical tree-based clustering of SOM weights
Start point: Vesanto & Alhoniemi, IEEE TNN'2000

Our Contributions: Using both the topology and the data density in clustering (neither was considered earlier)

Our preliminary results show significant improvement over the Vesanto & Alhoniemi algorithm.

The figures compare Vesanto and Alhoniemi's own clustering of the SOM weights learned from this 'clown' data set that they created by defining a distribution of 2-D points in the shape of a clown's eyes, nose, mouth and beard. The small pale blue dots represent the data points. The various color symbols (dots, crosses, open stars) indicate the location of neuron weights in the input space, and their cluster memberships. (In 2- or 3-dimensional data the SOM weights can be shown in the input space, like in this example. For higher dimensional data we have to do the visualization in the SOM lattice in the way we illustrated on the previous slides.) Note that on these figures the weight clusters are shown, not the corresponding clusters of data. However, it is fairly easy to see corresponding data clusters from the underlying pale blue data points.

Points to note:

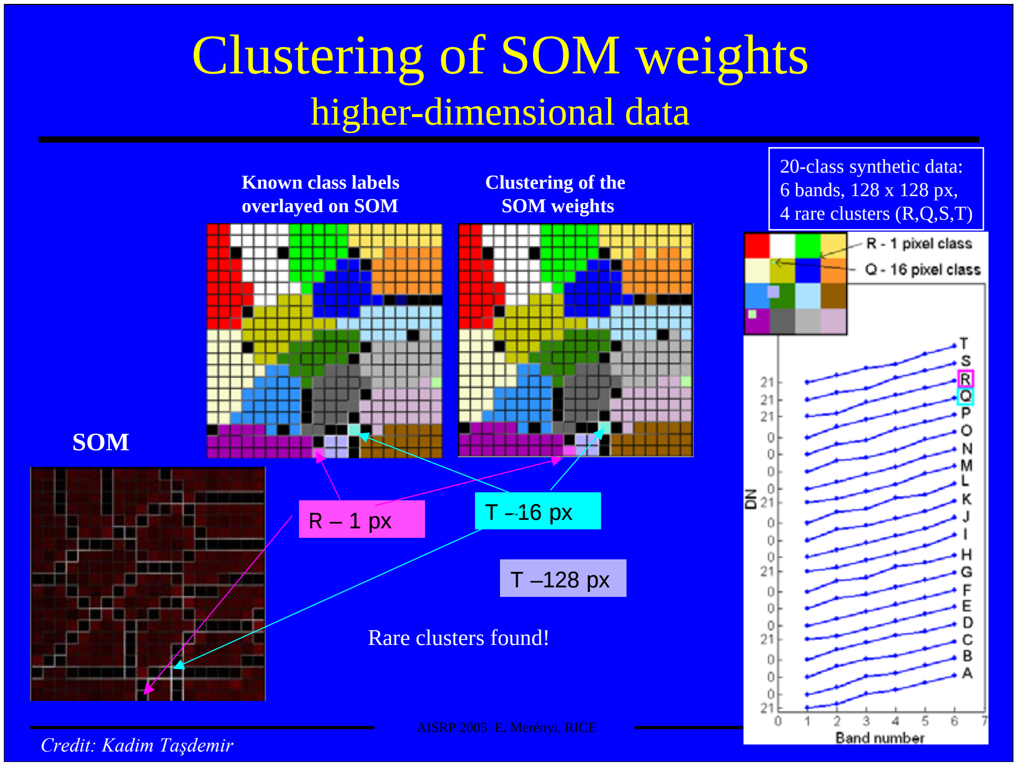
1) The two stray brown weights near the top of the Vesanto & Alhoniemi clustering (V & A from hereon) were clustered with the 'left eye' of the clown. These weights appear markedly fenced off in the U-matrix of the SOM in Vesanto & Alhoniemi, IEEE TNN'2000, yet – as the authors themselves remark – their algorithm lumps them into the brown cluster. This is the consequence of not considering the topology of the SOM. Our clustering captures them as singletons, each in its own cluster.

2) One brown and one black weight at either ends of the 'nose' of the clown are more correctly clustered into the nose by our algorithm (open green circles, overlayed on V & A's).

3) Part of the beard is clustered into the 'mouth' (red) by V & A. The same weights are – correctly – clustered into the beard by our method (open blue stars overlain on V & A's).

4) Perhaps the most important, a well defined cluster of data between the nose and the mouth, separated from both by only a small gap, was clustered into the nose by V & A. It is more correctly separated in our scheme (inset).

The correctness of the clusters is stated on the basis of inter- and intra-cluster distances.



A more challenging case: 6-dimensional data with 20 known classes (rightmost panel), in which the 20 classes have subtle differences among them. The set includes 4 rare classes to make it more challenging for automatic clustering.

Lower left: The learned SOM with fence and density structure.

Middle left: The known class labels overlain on the SOM.

Middle right: The clusters found by our automatic procedure.

Points to note:

- The identified clusters are very similar to the distribution of the known class labels (ground truth).
- The rare clusters were found.

Imperfections such as the difference in the distribution of yellow and orange classes are under current investigation.

An important SOM feature: Map Magnification

(Areal) magnification in neural maps is relationship between the *pdf* of the input data and the density of the SOM weights in input space.

$M(\mathbf{w}) \sim P(\mathbf{v})^{\alpha}$ where α is the *magnification exponent.*

Some useful properties:

$\alpha=1$ maximum entropy quantization (information theoretical optimum)

$\alpha = d/(d+2)$ minimum MSE distortion quantization where d is the dimension of the input space; $d = 1 \Rightarrow \alpha = 1/3$; $d = 2 \Rightarrow \alpha = 1/2$

$\alpha < 0$ enlarges response areas for low-frequency stimuli $\Rightarrow$ leads to better detection of small clusters

$\alpha = 2/3$ for Kohonen SOM --- not optimal in either min. distortion or max. entropy sense

$\alpha = 1$ is inherent in the Conscience algorithm, a variant of SOM, but it cannot produce other values of the magnification exponents

Magnification in neural maps is an important issue. Different magnification exponents (see above on slide) can help different data mining purposes, such as faithful detection of the pdf of data, or finding rare cluster (in the unknown data). Biological analog for negative magnification: neural maps in brains selectively magnify regions of interest for optimum information processing. Analog for alpha = 1 magnification: preservation of maximum information from layer to layer.

Magnification control

Bauer, Der and Hermann (BDH), 1996:

Modification of KSOM to allow α control

- by adaptively adjusting local learning rates, based on the winner node

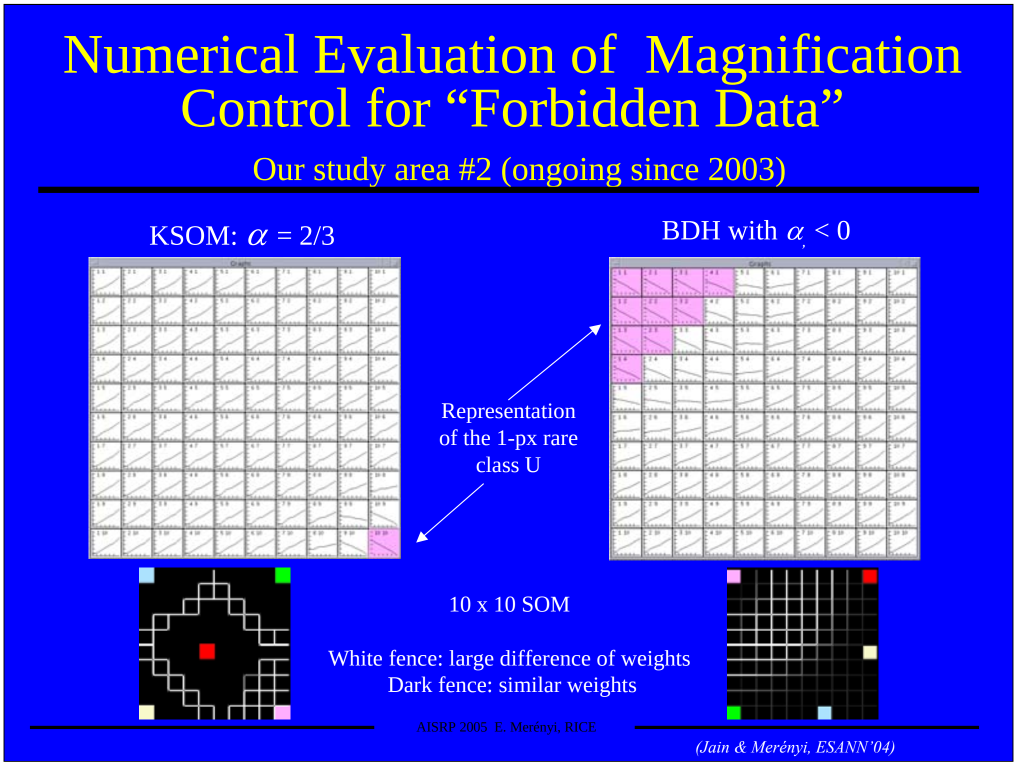
To induce $\alpha_{desired}$, learning rate should be

$$\varepsilon_r(t) = \varepsilon_0(t)\left[\frac{1}{\Delta t_r}\left(\frac{1}{|v - w_r|^d}\right)\right]^m$$

where $m = (3/2)\alpha_{desired} - 1$, d = "effective dimension"

But: theory supports very limited types of data …

The BDH algorithm provided a principled approach to controlling the magnification in SOMs. However, the thoretical proofs apply only to 1-D data or to n-D data where the n dimensions are statistically independent (the pdf separates to the marginals). We have been conducting numerical studies with "forbidden data" to scope the validity of the BDH magnification control. Expected benefits are largest for high-dimensional cases such as hyperspectral data.



We have been conducting numerical studies with "forbidden data" to scope the validity of the BDH magnification control. Expected benefits are largest for high-dimensional cases such as hyperspectral data (*Jain & Merényi, and Merényi & Jain, ESANN'04).* Negative magnification, in particular, is of interest because in principle it enhances the areal representation of very small clusters and thereby increases the chance of their discovery. Does it, however, work for data for which the theory cannot provide analytical support?

Left panel:

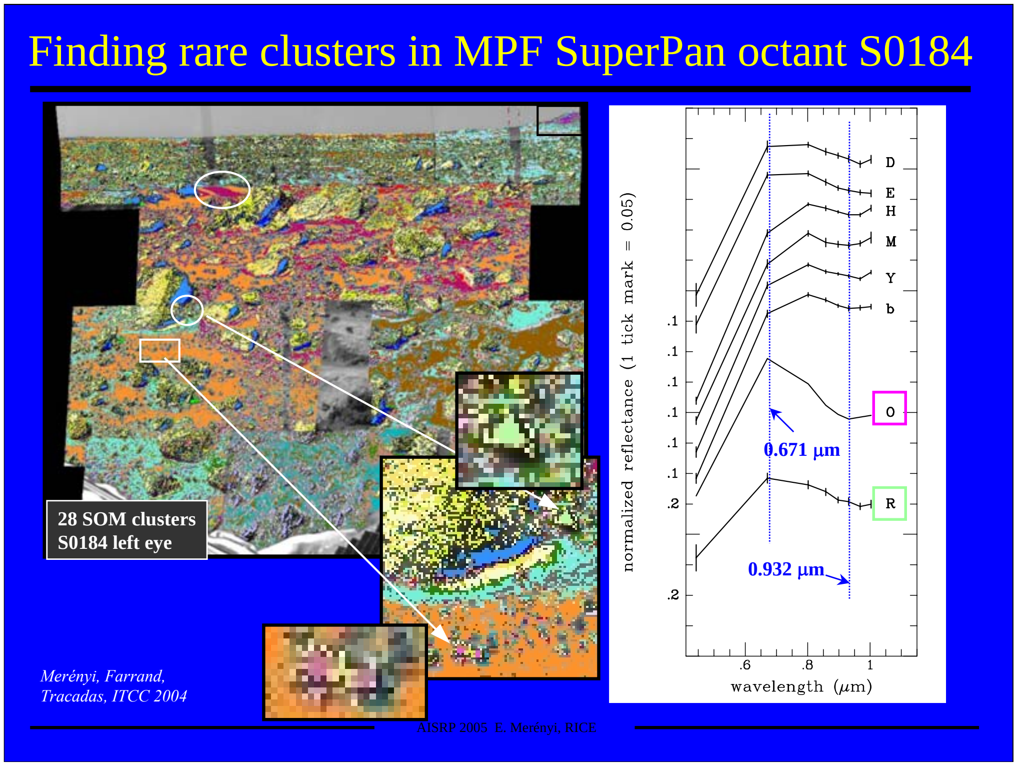
Top: Weight vectors learned by a 10 x 10 SOM using the basic Kohonen SOM (KSOM) learning. Only 1 PE represents the 1-pixel rare class U.

Bottom: Clusters identified in the map. The darker the fence between two PEs, the smaller the difference between the corresponding weights.

Right panel:

Top: The learned weights, using the BDH learning rule with negative magnification exponent. The rare class U is now represented by 10 weights!

Bottom: Clusters identified in the SOM.

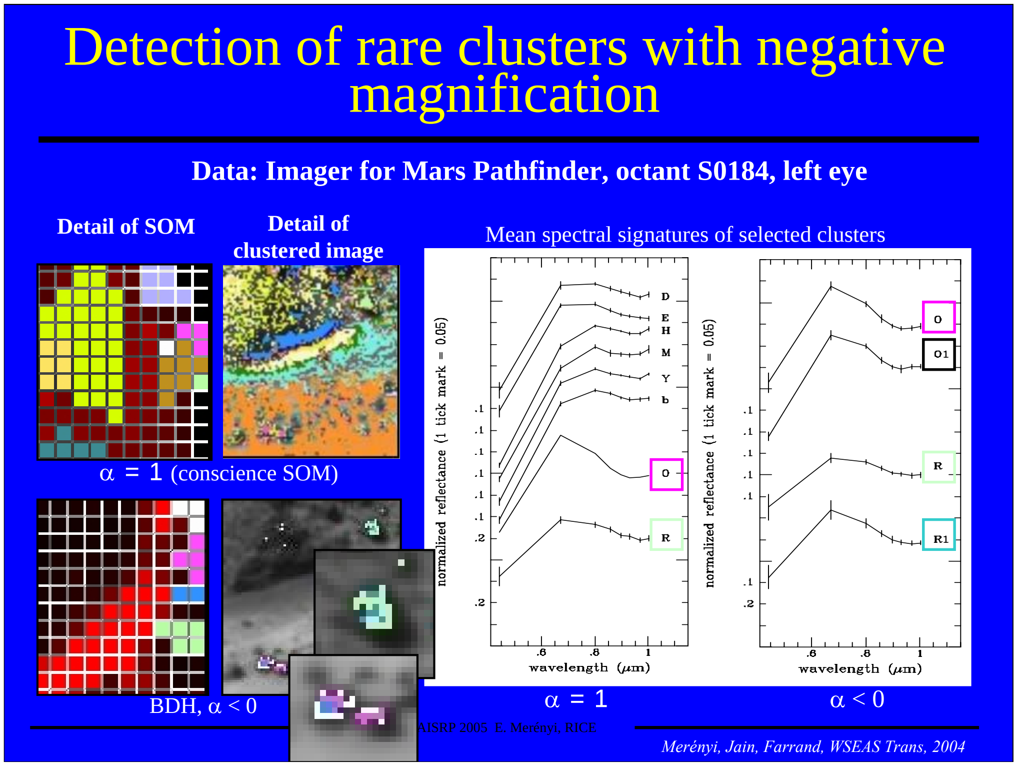


Clustering of one of the octants of the SUperPan panorama image taken by the Imager for Mars Pathfinder in 1997. This clustering, done with alpha = 1 (maximum entropy) SOM, identified the rare undifferentiated mineralogical type nicknamed “black rock” by geologists. In addition, the SOM found two subtypes within black rock occurrences: one (O) with an absorption shortward, and one (R) with an absorption longward of 0.95 microns. These subtypes can be interpreted as consistent with predominantly ortho- and clynopyroxene compositions.

The pink bow shaped feature (O) shown in the small lower rectangular inset is about 16 pixels in size, the triangle shaped pale green (R) feature (upper inset) comprises approvimately 30 pixels. There are not many more occurrences of O and R in this octant. The entire image is 1000 x 900 pixels, useful image size (without the data fallout areas and the air bag is ~ 600,000 pixels. The left eye of the IMP has 8 spectral bands.

In addition to clustering we produced supervised class maps of several of the SuperPan Octants (Farrand et al., LPSC 2005) after labeling the SOM-identified clusters. These classification maps are the first comprehensive spectral maps done from the SuperPan data, which turned out difficult. We use a hybrid ANN for supervised classification whose power is in containing an SOM as a hidden layer. Our earlier publications describe the approach. Slide 14 shows a sophisticated class map produced by such ANN.

Part of the difficulty with classification of the SuprPan panorama image is that it was stitched together from independently acquired “mosaics”. These mosaics were taken under varied conditions (instrumental, atmospheric, illumination, and perhaps others), and the differences remained in spite of three rounds of calibration efforts. As apparent from this cluster map, the SOM detected the subtle but consistent spectral differences that exist for the same cover type across mosaics. One example: the orange and brown clusters represent the same dust among the Martian rocks, but in the mosaic on the middle right (with obvious straight boundaries) the dust has a slightly different spectral signature than in the mosaic on the middle left. This could be used for improving the calibration.



This slide shows the effect of negative magnification for finding rare clusters in the same SuperPan octant as on the previous slide. Clearly, areal magnification occurred in the BDH SOM compared to Conscience SOM.

α Top two images:

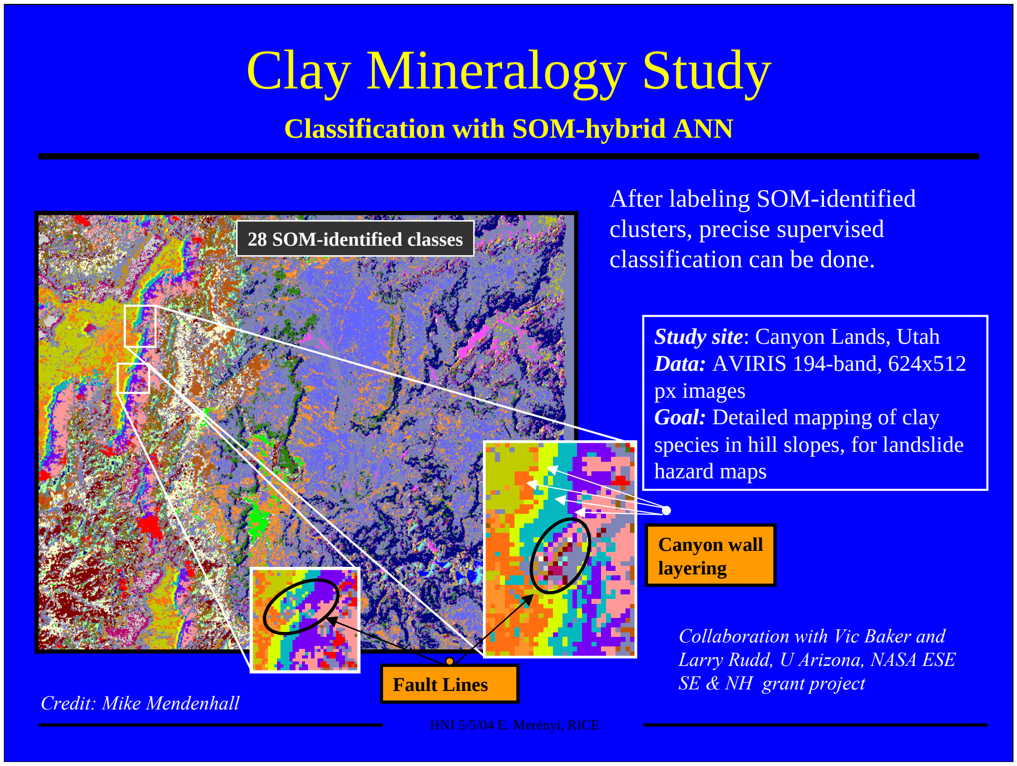
alpha = 1 case (maximum entropy quantization): The rare classes, pink (O) and light green, (R), are represented by 3 and 1 neurons, respectively.

2. Bottom two images:

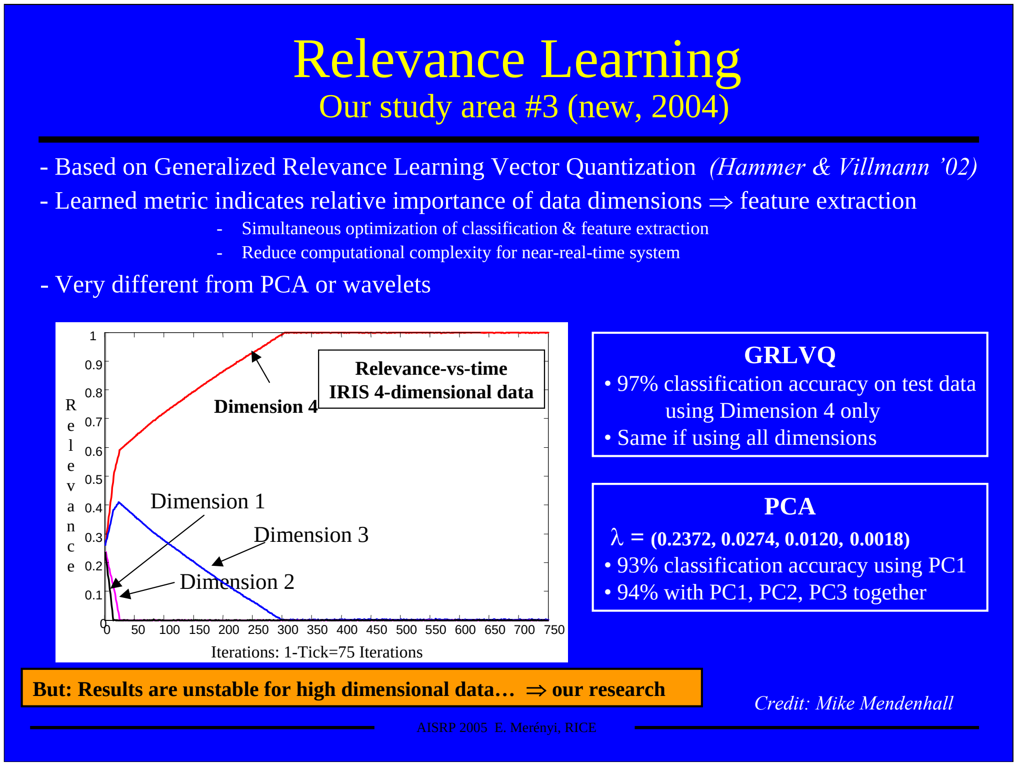
alpha < 0 case (negative magnification): The pink and white classes (O and O1) (7 neurons) together contain the former pink (O) class, the light green and blue (R and R1) together correspond to the former R class. Here, only the rare classes are shown on the spatial image, for clarity.

Subclusters were brought out within the tiny original O and R clusters! The spectral shapes (averages of the clusters, on the right panel of graphs) show that the substructure is justified.

White and blue colors are “recycled” for the BDH alpha < 0 case, they mean different spetcral clusters than in the Conscience SOM.



An example of detailed and precise supervised classification with SOM-hybrid ANN. The approach has been described in several earlier publications (e.g., Howell et al., JGR, 1994, Merényi et al., Icarus, 1997).



Relevance Learning

A) *What* -- Optimally identify the set of data components (features) needed to distinguish between known classes

B) *Why* -- Identify transform coefficients necessary for data compression

Remove superfluous data to maintain / increase classification performance

Reduce computational complexity for near-real-time systems

C) *Approach* -- Based on *Generalized Relevance Learning Vector Quantization* (GRLVQ) – Hammer & Villmann '02

Cast GRLVQ in a minimum classification error framework

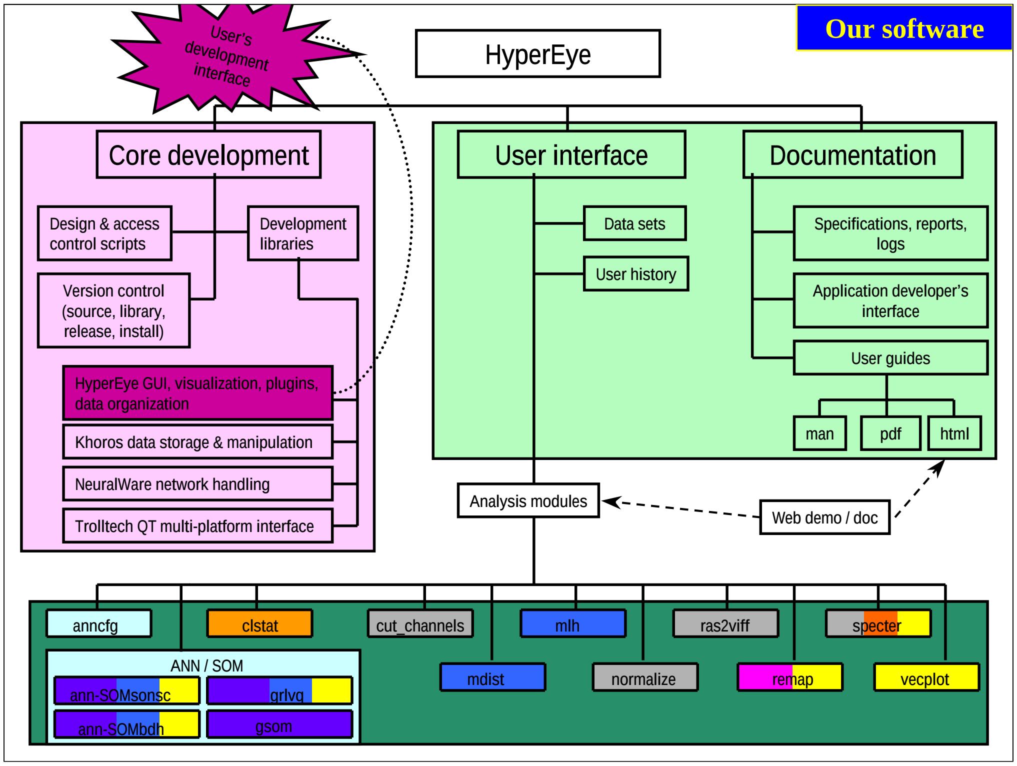
Apply GRLVQ to transform-based coding of hyperspectral data

This approach is very different from PCA or wavelets (for example), with which there has not been very much success in terms of preserving the discriminating information for classification.

The strength of GRLVQ over Principal Components Analysis, as we can assess from preliminary analyses, is illustrated above. While (for this relatively simple data set) GRLVQ singled out dimension 4 as 100% relevant and the other dimensions were evaluated as having 0 relevance, PCA distributes relevance a little differently in the transformed data. Yet, using the first three Principal Components together yields only 94% classification accuracy on test data, while classification using only the single dimension designated by GRLVQ produces 97% accuracy. 97% is about the best accuracy achievable for this data set because (as is known) it contains some mixed species.

Problem: the original Hammer & Villmann algorithm yields unstable results with hyperspectral data . We currently investigate stabilizing modifications by recasting the relevance framework from LVQ 2.1 to LVQ3 → emphasize in-class interactions.

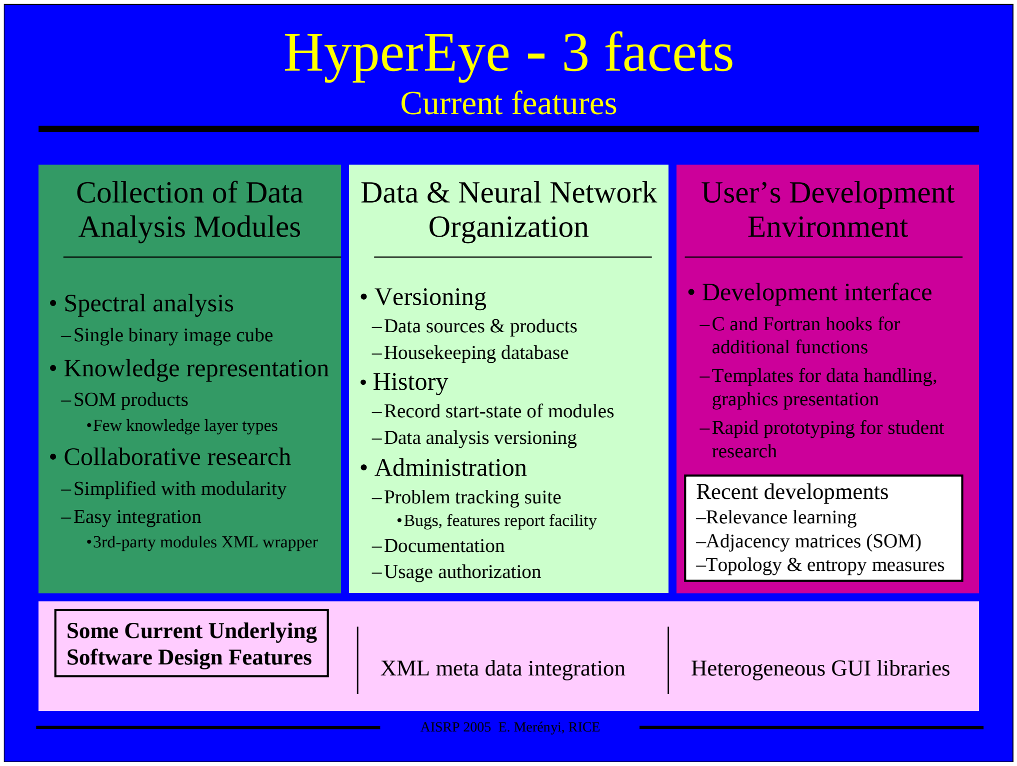
Results shown here are preliminary. Our present experiences are limited to a few data sets only.



Overview of the HyperEye system, our software environment for data analysis, algorithm research / design, and data organization.

The color coding of the large background boxes indicates major functional groups.

- Light pink = software design carpet including: source code control; XML meta data processing; history recording editor; underlying technology to wrap third-party libraries and implement our group's logic (for versioning, meta data, etc.); underlying technology to enable platform independence;
- Dark pink within light pink points out the development environment for users (as opposed to Application developer's environment on the right in light green): It greatly facilitates easy development of algorithms (new analysis modules) by researchers. This includes the development of new plugins (graphics, algorithm) to existing modules, for rapid prototyping of new subfunctionalities for larger modules.
- Dark green = collection of analysis modules. Color coding of analysis modules refers to their functional grouping. For example, light blue boxes are neural network related algorithms. Violet indicates SOM based neural modules; medium blue shows classification algorithms (neural or non-neural); grey codes preprocessing; yellow is graphics/visualization; orange is post-processing / evaluation of analysis results. Hot pink is cluster extraction (from a learned SOM).
- Light green = data and neural network organization: This is what the user interacts with: versioning of classification iterations, processing history, storage organization, administration (authenticate, docs, logs).



This is another view of the 3-faceted nature of HyperEye (top three panels), with examples of current features in each facets. Here we emphasize current limitations (instead of trying to be exhaustive), in order to show planned development on the next slide. For example,

In Analysis Modules:

- We currently handle single image cubes (no spatial relation is tracked between separate image cubes).

- The bullet under Knowledge representation of SOMs highlights a limit of the graphics, not a limit on analyses.

- Under Collaborative Research, 3rd-party module wrapper is through XML description to HyperEye describing what input file or command line format is/are needed to run the module.

In Data & Neural Network Organization:

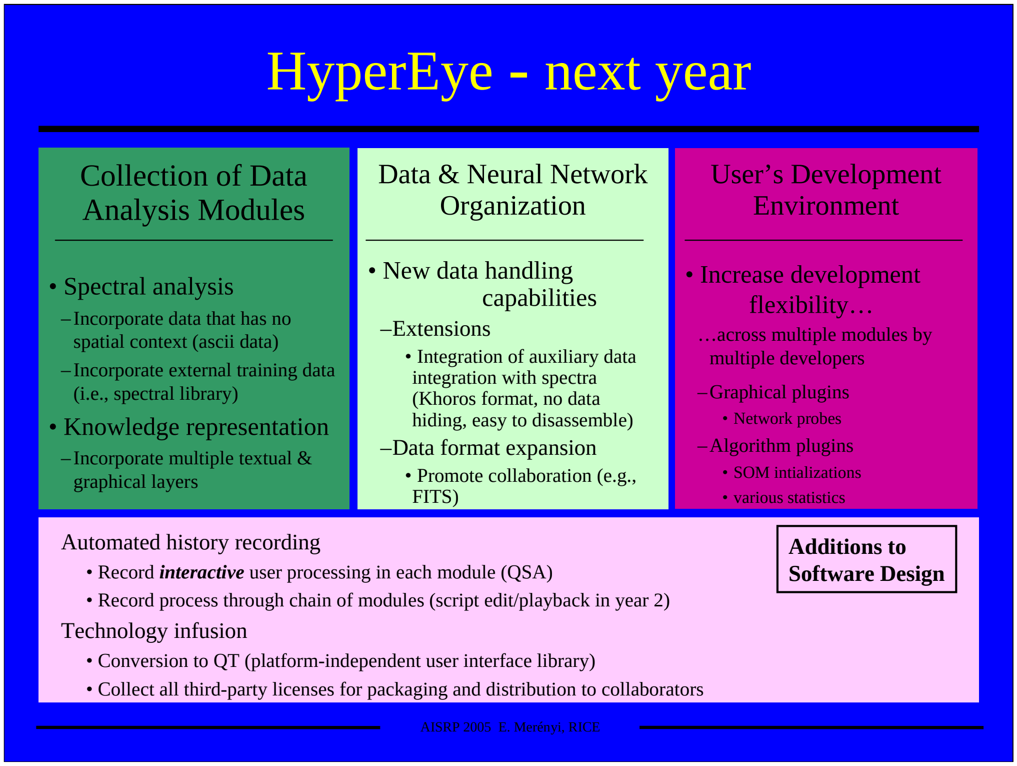
-Under Administration, usage authorization is now done through unix groups and mySQL user list (web bugs). This will get more elaborate with cross-platform data handling.

In the User’s Development Environment:

- Present GUI libraries are rather heterogeneous, reflecting layers of development history rather than organization for a more production oriented design.

From the Underlying Software Design (that supports all three facets):

- We point out the the integration of meta data is done by using XML, for a whole variety of things: module descriptions, meta data of source (scientific) data, history recording, etc. “Data” in this box is meant in a broad sense. For example, learned neural networks are viewed as data.



Some of the planned development for next year. Again, we emphasize planned changes within each of the three facets as well as in the underlying software design "carpet" rather than giving a comprehensive list.

In Development Environment: network probes is in reference to using Neural Ware's probes and instruments paradigm. Won't happen until year 2.

QSA: QT Scripting for Applications- a "visual basic" type of scripting.

Data Analysis for Space Missions, 2005

- Mars Exploration Rovers PanCam (6 spectral bands)
  - Also part of MDAP grant project (PI Bill Farrand, SSI)
- VIMS, Cassini (394 bands)
  - With Bob Brown, U AZ, VIMS PI
- Spectral study of icy volatiles (Pluto)
  - Collaboration with Eliot Young, PI, AISRP project
- Earth hyperspectral studies
  - Vic Baker, Larry Rudd, U AZ
- More IMP SuperPan
  - Try to reconcile calibration discrepancies based on cluster properties

A summary of the data analyses planned for 2005 in support of space missions.

Publications, 2004

http://www.ece.rice.edu/~erzsebet/publications.html

- **Merényi, E., (2004) Neural Maps for Precision Data Mining: Application to Planetary Spectral Images. Proc. Jordan Int'l Conference in Computer Science and Engineering, Al-Balqa University, Salt / Amman, Jordan, Oct 4 -7, 2004.**
- **Merényi, E., Jain, A., Farrand, W.H. (2004) Applications of SOM magnification to data mining. WSEAS Trans. on Systems, 3(5), July, 2004, pp 2122 - 2128.**
- **Jain, A., Merényi, E.(2004) Forbidden Magnification? I. Proc. 12th European Symposium on Artificial Neural Networks, ESANN'2004, Bruges, Belgium, 28-30 April, 2004, pp 51 - 56.**
- **Merényi, E., Jain, A., (2004) Forbidden Magnification? II. Proc. 12th European Symposium on Artificial Neural Networks, ESANN'2004, Bruges, Belgium, 28-30 April, 2004, pp 57 - 62.**
- **Merényi, E., Farrand, W.H., Tracadas, P. (2004) Mapping Surface Materials on Mars From Mars Pathfinder Spectral Images With HYPEREYE. Proc. International Conference on Information Technology (ITCC 2004), April 5-7, 2004, Las Vegas, NV, USA. vol II, pp 607 – 614.**
- **Farrand, W. H., Merényi, E., Murchie, S., Barnouin-Jha, O. (2005) Spectral Class Distinctions Observed in the MPF IMP SuperPan Using a Self-Organizing Map. Proc. 36th Lunar and Planetary Science Conference, Houston, Texas, March, 2005. Extended abstract.**
- **Rudd, L. and Merényi, E., The Use of AVIRIS Imagery to Assess Clay Mineralogy and Debris-Flow Potential in Cataract Canyon, Utah. Abstracts of the Conf. of the Geological Society of America, Denver, Colorado, Nov 9-11, 2004, Vol. 36, No. 5.**
- **W. H. Farrand and E. Merényi , Mapping Rock and Soil Units in the MPF IMP Superpan Using a Kohonen Self Organizing Map. Proc. 35th Lunar and Planetary Science Conference, Houston, Texas, March 15 – 19, 2004. Extended abstract.**

AISRP 2005 E. Merényi, RICE

Finally, our publications in 2004. Earlier and upcoming works are and will be listed / posted at the above URL.